

Please amend the claims as follows (this listing of claims replaces all prior versions):

1. (Currently Amended) A computer implemented method ~~of determining whether a set of nucleotides is within a first nucleic acid sequence, the method comprising:~~

receiving a first nucleic acid sequence;

generating a dictionary of words based on the first nucleic acid sequence such that the dictionary contains words that can be used to build the first nucleic acid sequence, the dictionary having a first number of words, each word having at least one nucleotide;

receiving a first and second nucleotide of a second nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combining said first and second nucleotide in sequence into a first set of nucleotides;

at a computer, comparing the first set of nucleotides to ~~a first nucleic acid sequence~~ the dictionary to determine whether the first set of nucleotides matches any word in the dictionary is within the first nucleic acid sequence; and

if the first set of nucleotides does not match any word in the dictionary is not within the first nucleic acid sequence, storing the first set of nucleotides as a new word in the dictionary unit in a database in one or more storage devices for the second nucleic acid sequence.

2. (Canceled)

3. (Currently Amended) The method of claim 1, wherein if the first set of nucleotides matches any word in the dictionary is within the first nucleic acid sequence, receiving a third nucleotide of the second nucleic acid sequence, the third nucleotide being a nucleotide after the second nucleotide.

4. (Original) The method of claim 3, further comprising: combining the first set of nucleotides with the third nucleotide to make a second sequential set.

5. (Currently Amended) The method of claim 4, further comprising: comparing the second set of nucleotides to the dictionary first nucleic acid sequence to determine whether the second set of sequential nucleotides matches any word in the dictionary ~~is within the first nucleic acid sequence.~~

6. (Currently Amended) The method of claim 5, wherein if the second set of nucleotides does not match any word in the dictionary ~~is not within the first nucleic acid sequence~~, storing said second set as a new word in the dictionary ~~unit in a database for the second nucleic acid sequence.~~

7. (Currently Amended) The method of claim 6, further comprising: determining the a second number of words in the dictionary after processing all the nucleotides in the second nucleic acid sequence ~~sum of all units stored for the second nucleic acid sequence.~~

8. (Currently Amended) The method of claim 7, further comprising: determining the difference between the second number and the first number ~~total number of units stored for the first nucleic acid sequence and the total number of units stored for the second nucleic acid sequence.~~

9. (Original) The method of claim 8, further comprising: utilizing the difference to determine the distance between the first nucleic acid sequence and the second nucleic acid sequence.

10. (Currently Amended) A non-transitory computer readable storage medium comprising instructions that when executed by a machine, causes the machine to perform:
identify a first nucleic acid sequence;
generate a dictionary of words based on the first nucleic acid sequence such that the dictionary contains words that can be used to build the first nucleic acid sequence, the dictionary having a first number of words, each word having at least one nucleotide;

receive a first and second nucleotide of a second nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combine the first and second nucleotide in sequence into a first set of nucleotides;

compare the first set of nucleotides to the dictionary first nucleic acid sequence to determine whether the first set of nucleotides matches any word in the dictionary is within the first nucleic acid sequence; and

if the first set of nucleotides does not match any word in the dictionary is not within the first nucleic acid sequence, store the first set of nucleotides as a new word in the dictionary unit in a database for the second nucleic acid sequence.

11. (Currently Amended) A computer implemented method of creating a database of nucleotide units for a first nucleic acid sequence, the method comprising:

receiving a first nucleotide of a first nucleic acid sequence;

at a computer, determining whether the first nucleotide has been stored in a database in one or more storage devices as a unit for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence;

if the first nucleotide has not been stored in the database separately from the first nucleic acid sequence, storing the first nucleotide as an individual unit for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence;

if the first nucleotide has been stored in the database separately from the first nucleic acid sequence, receiving a second nucleotide of the first nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combining the first and second nucleotides into a sequential set;

at the computer, determining whether the sequential set has been stored in the database as a unit for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence; and

if the sequential set has not been stored in the database separately from the first nucleic acid sequence, storing the sequential set as a unit in the database for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence.

12 - 14. (Canceled)

15. (Previously Presented) The method of claim 11, wherein if the sequential set has been stored, receiving a third nucleotide of the first nucleic acid sequence, the third nucleotide being the next sequential nucleotide after the second nucleotide.

16. (Canceled)

17. (Currently Amended) The method of claim 11, further comprising: determining the sum of all units stored for the first nucleic acid sequence.

18. (Currently Amended) A non-transitory computer readable ~~storage~~ medium comprising instructions that when executed by ~~a machine~~ one or more machines causes the one or more machines ~~machine~~ to:

receive a first nucleotide of a first nucleic acid sequence;

determine whether the first nucleotide has been stored in a database in one or more storage devices as a unit for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence;

if the first nucleotide has not been stored in the database separately from the first nucleic acid sequence, store the first nucleotide as an individual unit for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence;

if the first nucleotide has been stored in the database separately from the first nucleic acid sequence, receive a second nucleotide of the first nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combine the first and second nucleotides into a sequential set;

determine whether the sequential set has been stored in the database as a unit for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence; and

if the sequential set has not been stored in the database separately from the first nucleic acid sequence, store the sequential set as a unit in the database for the first nucleic acid sequence, the unit being separate from the first nucleic acid sequence.

19. (Currently Amended) A system for determining a distance between a first nucleic acid sequence and a second nucleic acid sequence, the system comprising one or more storage units and one or more data processors executing instructions to implement:

~~a receiving component for receiving a first nucleic acid sequence;~~

generating a dictionary of words based on the first nucleic acid sequence such that the dictionary contains words that can be used to build the first nucleic acid sequence, the dictionary having a first number of words, each word having at least one nucleotide;

receiving a first and a second nucleotide of a second nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

~~a combining component for combining said first and second nucleotide in sequence into a first set of nucleotides;~~

~~a comparing component for comparing the first set of nucleotides to the dictionary a first nucleic acid sequence to determine whether the first set of sequential nucleotides matches any word in the dictionary is within the first nucleic acid sequence;~~

~~a storing component for storing said first set as a new word in the dictionary unit in a database for the second nucleic acid sequence if the first set of nucleotides does not match any word in the dictionary is not within the first nucleic acid sequence.~~

20. (Canceled)

21. (Currently Amended) The system of claim 19, ~~comprising a second wherein the receiving module for system further implements~~ receiving a third nucleotide of the second nucleic acid sequence if it is determined that the first set of nucleotides matches a word in the dictionary is within the first nucleic acid sequence.

22. (Currently Amended) A computer-implemented method of determining the distance between two nucleic acid sequences, the method comprising:

determining the number of words in a first nucleic acid sequence;

combining the first sequence with a second nucleic acid sequence to make a combined nucleic acid sequence;

determining the number of words in the combined nucleic acid sequence; and

determining at one or more computers, a number representing the difference between the number of words in the combined nucleic acid sequence and the first nucleic acid sequence to determine the distance between the first nucleic acid sequence and the second nucleic acid sequence.

23. (Currently Amended) A non-transitory computer readable storage medium comprising instructions that when executed by ~~a machine~~ one or more computers cause the one or more computers ~~machine~~ to:

determine the number of words in a first nucleic acid sequence;

combine the first sequence with a second nucleic acid sequence to make a combined nucleic acid sequence;

determine the number of words in the combined nucleic acid sequence; and

determine the difference between the number of words in the combined nucleic acid sequence and the first nucleic acid sequence to determine a distance between the first nucleic acid sequence and the second nucleic acid sequence.

24. (Canceled)

25. (Canceled)

26. (Canceled)

27. (Canceled)

28. (Canceled)

29. (Canceled)

30. (Canceled)

31. (Currently Amended) The method of claim [[1]] 32, comprising in which determining a distance between the first and second nucleic acid sequences comprises determining the distance between the first and second nucleic acid sequences based on a distance

measure for nucleic acid sequences that satisfies the triangle inequality, such that a distance between the first and second nucleic acid sequences is no greater than a sum of a first distance between the first nucleic acid sequence and a third nucleic acid sequence, and a second distance between the second and third nucleic acid sequences.

32. (New) A computer implemented method of determining a distance between a first nucleic acid sequence and a second nucleic acid sequence, the method comprising:

determining a first number representing the number of words in a first dictionary that can be used to build a first nucleic acid sequence, each word comprising at least one nucleotide;

determining a second number representing the number of words in a second dictionary that can be used to build a second nucleic acid sequence,;

determining a third number representing the number of words in a third dictionary that can be used to build a first combined nucleic acid sequence comprising the second nucleic acid sequence appended to the first nucleic acid sequence;

determining a fourth number representing the number of words in a fourth dictionary that can be used to build a second combined nucleic acid sequence comprising the first nucleic acid sequence appended to the second nucleic acid sequence; and

determining, at one or more computers, a distance between the first and second nucleic acid sequences based on the first number, the second number, the third number, and the fourth number.

33. (New) The method of claim 32, comprising:

determining a first difference between the third number and the first number;

determining a second difference between the fourth number and the second number; and

determining the distance between the first and second nucleic acid sequences based on a maximum of the first difference and the second difference.

34. (New) The method of claim 33, comprising determining a normalized distance between the first and second nucleic acid sequences based on the maximum of the first

difference and the second difference, divided by a maximum of the first number and the second number.

35. (New) The method of claim 32, comprising
determining a first difference between the third number and the first number;
determining a second difference between the fourth number and the second number; and
determining the distance between the first and second nucleic acid sequences based on a sum of the first difference and the second difference.

36. (New) The method of claim 35, comprising determining a normalized distance based on the sum of the first difference and the second difference, the sum being divided by the third number.

37. (New) The method of claim 35, comprising determining a normalized distance based on the sum of the first difference and the second difference, the sum being divided by an average of the third number and the fourth number.

38. (New) The method of claim 32 in which generating the first dictionary comprises:
receiving a first and second nucleotide of the first nucleic acid sequence, the second nucleotide being after the first nucleotide in the first nucleic acid sequence;
combining the first and second nucleotides in sequence into a first set of nucleotides; and
adding the first set of nucleotides as a word into the first dictionary if the first set of nucleotides is not already in the first dictionary.

39. (New) The method of claim 38 in which if the first set of nucleotides is already in the first dictionary, receiving a third nucleotide of the first nucleic acid sequence, the third nucleotide being after the second nucleotide,
combining the first set of nucleotides with the third nucleotide to make a second set of nucleotides, and

adding the second set of nucleotides as a word into the first dictionary if the second set of nucleotides is not already in the first dictionary.